Theoretical Computer Science 58 (1988) 103–128
North-Holland

FINITE BIPREFIX SETS OF PATHS IN A GRAPH *

Clelia DE FELICE

Dipartimento di Matematica e Applicazioni, Università di Napoli, 80100 Naples, Italy, and LITP, Université de Paris VI et VII, 75251 Paris Cedex, France

Abstract. The main results of the combinatorial theory of maximal biprefix codes of words (Césari, Perrin, Schützenberger) are extended to the codes of paths in a graph in this paper: degree and decoding of double-infinite paths, finiteness of codes of a given degree, the Césari–Schützenberger algorithm, derivation and integration of codes will be discussed.

1. Introduction

The theory of *variable-length codes* (i.e., the bases of free submonoids of a free monoid), born in Shannon's early works on information transmission, has been developed in an algebraic direction by Schützenberger and his school, and in the last twenty years has become a considerable part of theoretical computer science.

Techniques and results of automata and finite monoid theory, of formal power series and language theory proved to be powerful tools for studying codes. Moreover, results and problems in these fields can be seen and formulated as results and problems of the theory of codes. For a complete survey of the theory of codes see [1].

One of the problems of this theory is the study of several families of codes. The structure of *maximal biprefix codes* in particular has been intensively investigated by Césari, Perrin and Schützenberger [1, 3, 4, 6, 9].

Recently, Reutenauer [8] has provided the bases of an extension of the theory of codes to the case of paths in directed graphs, introducing the concept of *code of paths*. This is the counterpart of the free monoids of Tilson's theory of semigroupoids [10].

The aim of this paper is to extend the main results of the combinatorial theory of maximal biprefix codes to the codes of paths in a graph. *Degree* and decoding of double-infinite paths, finiteness of codes of a given degree, the *Césari–Schützenberger algorithm*, *derivation* and *integration* of codes will be discussed.

This direction of research is motivated not only by the pleasure of generalization, but was, at least at the outset, also an attempt to study some local properties of codes (and more generally of automata) limiting the conditions on the sequencing of the letters. It is simply an investigation into some properties of specific formal languages.

* This research was partially supported by C.N.R. (Consiglio Nazionale delle Ricerche, Roma, Italy).

As to the relationship between the classic theory of codes and its extension to the paths in a graph, it must be pointed out that some concept (e.g., the notion of biprefix code and maximal biprefix code) are easily transferred from one to the other.

Other concepts (e.g., completion relative to a vertex, cf. Section 6) are specific to the theory of codes of paths. Their introduction is due to the fact that several results for codes of words must be fit for codes of paths (see Lemmas 6.6, 6.7).

Finally, for some concepts (e.g., the probabilistic interpretation of the notion of code and of maximal finite code) it is not yet known whether they can be transferred to the case of the codes of paths. Obviously, a proof must be found which is different from the classic one to extend the results of the theory of codes of words which make use of these concepts.

In this paper we will prove only those results whose demonstrations differ from that of the classic theory of codes, and we will enounce only those whose demonstrations are an evident generalization of that of the classic theory.

We consider a *strongly connected (directed) graph* G. A *biprefix code* over G is a set C of paths of G such that no path in C is either a left factor or a right factor of another path in C. A finite biprefix code C is termed *complete* if any sufficiently long path in the graph admits exactly one left and one right factor in C.

We come again to the classical case of the biprefix codes of words [1, 3, 5, 7] when we limit ourselves to the graphs with one vertex.

We begin by showing that the number of decodings of any double-infinite path by a finite complete biprefix code is an integer depending only on C (Theorem 3.2): this number will be called the *degree* d of C.

An example of a finite complete biprefix code over the graph as shown in Fig. 1 is the code

$$C = \{aa, abc, b, abd, dc, ca, cbc, dd, cbd\}.$$

C can be seen to be of degree 2, a case that cannot take place with the codes of words.

If the graph G has a loop (i.e., a closed path of length 1) a, then a^d is always in C (Theorem 4.3) and more generally, for any closed path c of G, there exists a power of c which is the product of paths of C (Proposition 4.2).

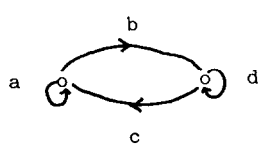

Fig. 1.

We show that only a finite number of codes of a given degree (Theorem 5.6) exists as in the case of words. Moreover, we extend the Césari-Schützenberger algorithm to the codes of paths: any code of degree d can be obtained by a finite number of internal transformations starting with the homogeneous code of the same degree d (i.e., the set of paths of length d) (Theorems 6.5 and 6.8).

We extend the definition of derivation of a code (according to Césari): the derivative of a code of degree d is a code of degree $d-1$ (Theorem 7.3) and we will show that any code of degree $d-1$ is the derivative of a code of degree d (Theorem 8.12).

2. Definitions

Let $G=(V, A)$ be a *directed graph*, where V is the finite set of *vertices* and A the finite set of *arrows*. A^* is the set of the *paths* in G; 1_v, for any $v \in V$, the empty path from v to v; A^+ the set of the nonempty paths in G. In fact, (V, A^*) is the free category generated by G.

In the following we suppose that G is a *strongly connected* graph (i.e., for any $(v, t) \in V^2$ there exists a path from v to t).

For any $c \in A^*$, let us denote i_c, t_c respectively the initial and the terminal vertex of c, and $|c|$ the length of c (i.e., the number of arrows composing it). Let c_1, c_2, $c_3 \in A^*$ be paths such that $c = c_1c_2c_3 \in A^*$. Then c_2 is a *factor* (a *proper factor* if c_1 and c_3 are nonempty) of c; c_3 is a *suffix* (a *proper suffix* if c_1c_2 is nonempty) of c; and c_1 is a *prefix* (a *proper prefix* if c_2c_3 is nonempty) of c.

The partial product of paths naturally defines a partial product in the set $\mathcal{P}(A^*)$ of the subsets of A^*, i.e., for all subsets X, Y of A^*,

$$XY = \{w \in A^* \mid \exists x \in X, y \in Y: w = xy\}.$$

Moreover, for any $C \subseteq A^*$, let C^* be the set of paths obtained by concatenation of paths of C, including the empty paths of G. We can extend the definitions of the theory of codes to the sets of paths [1, 5, 7, 8].

A set $C \subseteq A^*$ is a *code* if for any $c_1, \ldots, c_h, c'_1, \ldots, c'_k$ in C such that $c_1 \ldots c_h \in A^*$ we have

$$c_1 \ldots c_h = c'_1 \ldots c'_k$$
$$\Rightarrow h = k \text{ and } \forall i \in \{1, \ldots, h\}: c'_i = c_i.$$

A set $C \subseteq A^+$ is a *prefix* (*suffix*) code if

$$CA^+ \cap C = \emptyset, \qquad (A^+C \cap C = \emptyset).$$

A code is *biprefix* if it is both a prefix code and a suffix code. *In the following the prefix, suffix and biprefix codes considered will be finite subsets of* A^*.

A *prefix* (*suffix*) code is *complete* if

$$\forall c \in A^*: \quad cA^* \cap C^* \neq \emptyset, \qquad (A^*c \cap C^* \neq \emptyset).$$

This condition can be seen to be equivalent to the maximality of C as a prefix (suffix) code; i.e., for any prefix (suffix) code C' we have

$$C \subseteq C' \Rightarrow C = C'.$$

A *biprefix* code is *complete* if it is both a prefix complete code and a suffix complete code.

Let us denote A^d the homogeneous code (of degree d)

$$A^d = \{c \in A^* \mid |c| = d\}.$$

As G is a strongly connected graph, A^d is a complete biprefix code.

3. Degree

This section presents the notion of *degree* of a finite complete biprefix code. First, some definitions.

A *double-infinite path* [8] is a mapping $z : \mathbb{Z} \to A$ such that for any $i \in \mathbb{Z}$ the product $z(i)z(i+1)$ is defined in G. In fact, any double-infinite path is an infinite sequence of consecutive arrows $(z(i))_{i \in \mathbb{Z}}$.

For any $h, k \in \mathbb{Z}$, $h < k$, $z[h, k]$ is the path $z(h) \dots z(k)$. A path c is a *factor* of z if there are $h, k \in \mathbb{Z}$, $h < k$, such that $z[h, k] = c$.

Let $C \subseteq A^*$ be a biprefix code. For any $t \in \mathbb{Z}$ let τ_t be the translation of $\mathbb{Z}$ defined by

$$\forall i \in \mathbb{Z}: \quad \tau_t(i) = i + t.$$

In the set of the strictly increasing mappings μ from $\mathbb{Z}$ to $\mathbb{Z}$ such that

$$\forall i \in \mathbb{Z}: \quad z(\mu(i))z(\mu(i)+1) \dots z(\mu(i+1)-1) \in C,$$

let us define the following equivalence relation $\equiv$:

$$\mu \equiv \mu' \Leftrightarrow \exists t \in \mathbb{Z}: \mu = \mu' \circ \tau_t.$$

A *decoding* of z in C is an equivalence class mod $\equiv$. In fact, a decoding is a factorization of z into elements of C. Thus in the following we identify decoding and factorization. For instance, if C is the code $\{aa, abc, b, abd, dc, ca, cbc, dd, cbd\}$ on the graph shown in Fig. 1 above, and if we consider the double-infinite path

$$\dots abcabcabcabc \dots,$$

then two decodings are admitted which are shown in Fig. 2.

A *point* of a double-infinite path z is a factorization of z as a product of two paths. Formally, a point of z is a pair (x_j, y_j), where $j \in \mathbb{Z}$, x_j is the restriction of z to the set $\{i \in \mathbb{Z} \mid i < j\}$ and y_j is the restriction of z to the set $\{i \in \mathbb{Z} \mid i \geqslant j\}$. We shall also say that j is a point of z.

A *decoding* μ of z in C *passes through the point* (x_j, y_j) of z if there exists an $i \in \mathbb{Z}$ such that $\mu(i) = j$, i.e., if a cut of the factorization of z into elements of C is between $z(j-1)$ and $z(j)$ (see Fig. 3).

Fig. 2.

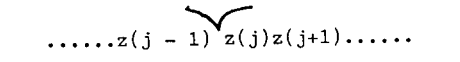

Fig. 3.

Lemma 3.1. *Let C be a finite complete biprefix code and z a double-infinite path. For each point of z there is one and only one decoding which passes through this point.*

Proof. Since C is a complete for any $j \in \mathbb{Z}$, there is one path of C on the left of $z(j)$ and another on the right of $z(j-1)$, i.e.,

$$\forall j \in \mathbb{Z}, \exists (h, k) \in N^2:$$
$$z(j)z(j+1) \dots z(j+h), z(j-k-1) \dots z(j-1) \in C.$$

Since C is biprefix, this is equivalent to saying that for any $j \in \mathbb{Z}$ there is a factorization of z into elements of C passing between $z(j-1)$ and $z(j)$, i.e., that for each point j of z there is one decoding which passes through it.

By contradiction, let μ, μ' be two different decodings of z passing through the same point j. Therefore, we can suppose $\mu(j) = \mu'(j) = j$. Then there exists an $h \in \mathbb{Z}$ with $\mu(h) = \mu'(h) = h$ and such that either $\mu(h+1) \neq \mu'(h+1)$ or $\mu(h-1) \neq \mu'(h-1)$, contradicting the hypothesis that C is a biprefix code. □

Theorem 3.2. *Let C be a finite complete biprefix code. Then any double-infinite path has the same number d of decodings in C.*

Proof. First we prove that any double-infinite path z has a finite number of decodings. Let $j, k \in \mathbb{N}$ such that $|z(j) \dots z(j+k)|$ is greater than or equal to the maximal length of the paths of C. By definition of j and k any decoding of z passes through one of the points $\{j, \dots, j+k\}$ of z. Then, by Lemma 3.1, there exists $k+1$ decodings of z at most.

Let z_1, z_2 be two double-infinite paths.

(1) Suppose that property (P) holds for z_1 and z_2:

$$(P) \quad \exists (j, k) \in \mathbb{Z}: \quad z_1(j)z_2(k) \text{ is defined in } G.$$

Then z_1 and z_2 have the same number of decodings. Indeed, let z_3 be the double-infinite path

$$\dots z_1(j-1)z_1(j)z_2(k)z_2(k+1) \dots$$

obtained by attaching the part of z_1 on the left of $z_1(j+1)$ to the part of z_2 on the right of $z_2(k-1)$.

Let μ be a decoding of z_1. This decoding passes through some point $p \leqslant j$ of z_1. By Lemma 3.1, there exists a decoding ν of z_3 equal to μ in the part of z_3 on the left of $z_2(k)$. By the same lemma, there exists one and only one decoding μ' of z_2 equal to ν in the part of z_3 on the right of $z_1(j)$. Moreover, if we take a decoding of z_2, then by a symmetric argument we can associate to it one and only one decoding

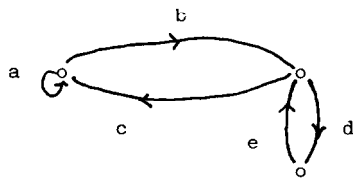

Fig. 4.

Fig. 5.

of z_1. Then the mapping $\mu \to \mu'$ is a bijection from the set of decodings of z_1 on the set of decodings of z_2. Then z_1 and z_2 have the same number of decodings.

(2) Now let z_1, z_2 be two double-infinite paths. Since G is strongly connected, for any $j, k \in \mathbb{Z}$ there is a $t \in A^*$ such that $z_1(j)tz_2(k)$ is defined in G. Let z_3 be the double-infinite path

$$\ldots z_1(j-2)z_1(j-1)z_1(j)tz_2(k)z_2(k+1)\ldots$$

Property (P) holds for the pairs (z_1, z_3) and (z_3, z_2). Then, by (1), z_1 and z_3 have the same number of decodings in C, as do z_3 and z_2; i.e., z_1 and z_2 have the same number of decodings in C. □

This proposition justifies the following definition.

Definition. Let C be a finite complete biprefix code. The *degree* of C is the number d of decodings in C of any double-infinite path.

For instance, in the preceding example C has degree 2. Let $B = \{a, b, c, d, e\}$ and let $D = \{aa, ab, bc, bd, ca, cb, e, ded, dec\}$ be the complete biprefix code on the graph shown in Fig. 4. D has degree 2. Indeed, the double-infinite path

$$\ldots abdecabdecabdec\ldots$$

has two decodings in D as shown in Fig. 5.

4. Biprefix codes and closed paths

We prove some consequences of the results of the preceding paragraph. A *loop* is a closed path of length 1.

Remark 4.1. If C is a code of paths and c is a closed path, there exists at most one power of c in C. Indeed let $i, j \in \mathbb{N}^2$ be such that $c^i, c^j \in C$. Then we have $(c^i)(c^j) = (c^j)(c^i)$ and C is a code if and only if $i = j$.

Proposition 4.2. *Let C be a finite biprefix code of paths. C is complete if and only if for any closed path $c \in A^*$ there exists a power of c in C^*.*

Proof. Let us suppose that for any closed path c there exists a power of c in C^* and let $c' \in A^+$. Since G is strongly connected, there exists a $t \in A^*$ such that $c'tc' \in A^*$. Then $c't$ and tc' are closed paths and, by hypothesis, there exist $m, n \in \mathbb{N}$ such that

$$(c't)^m = c'(t(c't)^{m-1}) \in C^*, \qquad (tc')^n = ((tc')^{n-1}t)c' \in C^*.$$

Therefore, C is a complete biprefix code.

Conversely, let C be a finite complete biprefix code and $a_1, \ldots, a_n \in A$ such that $c = a_1 \ldots a_n$ is a closed path. Let us consider the double-infinite path z defined by $z = \ldots ccc \ldots$. By Lemma 3.1, for each point of z dividing two consecutive occurrences of c there exists a decoding passing through it. There is an infinite number of these points but, by Theorem 3.2, there is only a finite number of decodings. Then there exists a decoding passing through two of these points, i.e., there is a $k \in \mathbb{N}$ such that $c^k \in C^*$. □

Theorem 4.3. *Let C be a finite complete biprefix code of degree d. For any loop $a \in A$ we have $a^d \in C$.*

Proof. Let a be a loop and let us consider the double-infinite path

$$z = \ldots aaaa \ldots .$$

By Proposition 4.2 and Remark 4.1, there is one and only one power a^k of a in C. Then a factorization of z is a decoding if and only if it factorizes z in the product of paths all equal to a^k. Therefore, z has k decodings in C and, by Theorem 3.2, we have $k = d$, i.e., $a^d \in C$. □

5. The set of finite complete biprefix codes of a given degree is finite

The aim of this section is to prove that there exists only a finite number of finite complete biprefix codes of a given degree d (Theorem 5.6). Therefore, we extend some definitions on words to the area of paths.

Definition. A *quasi-power of order* 0 is a nonempty path. A quasi-power of order $n+1$ is a path c having the form xyz where x is a quasi-power of order n. As in the case of words, by a similar proof [2] we have the following proposition.

Proposition 5.1. *For any $n \in \mathbb{N}$ there exists a $k(n) \in \mathbb{N}$ such that any path of length at least $k(n)$ has a quasi-power of order n as factor.*

Let C be a finite complete biprefix code of paths. Let us denote by S the set of the proper suffixes of the elements of C, and P the set of the proper prefixes of the elements of C.

A *C-interpretation* of a path c is a triple (s, x, p) with $s \in S$, $x \in C^*$, $p \in P$ and such that $c = sxp$. Since C is a complete biprefix code, any path has at least one *C-interpretation.*

Let c be a path. A *point* of c is a pair (c_1, c_2) of paths such that $c = c_1 c_2$. We say that a C-interpretation (s, x, p) of c *passes through the point* (c_1, c_2) if there exists a factorization of x, $x = x_1 x_2$, with $x_1, x_2 \in C^*$ and such that

$$c_1 = sx_1, \qquad c_2 = x_2 p.$$

Let $H(C)$ be the set of the proper factors of the elements of C, and λ the mapping such that $\lambda(c)$ is the number of C-interpretations of c for any $c \in A^*$. As in the case of words, by a similar argument [1] we have the following proposition.

Proposition 5.2. *Let C be a finite complete biprefix code and c a path. Then for each point of c there is one and only one C-interpretation passing through this point.*

Proposition 5.3. *For any finite complete biprefix code C and for any $c \in A^*$, $\lambda(c)$ is equal to the number of the suffixes of c belonging to P and is equal to the number of the prefixes of c belonging to S.*

Proof. If (s, x, p) is an interpretation of C, then $p \in P$ is a suffix of c.

Conversely, let c_2 be a suffix of c belonging to P. Then there exists a $c_1 \in A^*$ such that $c = c_1 c_2$.

By Proposition 5.2, there exists one and only one C-interpretation passing through the point (c_1, c_2) of c. Then there exist an $x \in C^*$ and a $p \in P$ such that $c_2 = xp$. But $c_2 \in P$ implies that c_2 has no prefix in C, therefore, x is an empty path.

Let $p \in P$ be such that (s, x, p), (s_1, x_1, p) are C-interpretations of c. By Proposition 5.2, we have

$$(s, x, p) = (s_1, x_1, p).$$

Then the mapping $(s, x, p) \to p$ is a bijection from the set of C-interpretations of c on the set of the suffixes of c belonging to P.

The second part of the statement can be proven in a similar way. □

As in the case of the codes of words, by an analogous argument [1] we have the next proposition.

Proposition 5.4. *Let C be a finite complete biprefix code of degree d. Then*:

(i) *for any $r, c, q \in A^*$ such that $rcq \in A^*$ we have $\lambda(rcq) \geqslant \lambda(c)$;*

(ii) *λ is bounded, constant on the set of paths not belonging to $H(C)$ and, for any $c' \notin H(C)$, we have*

$$\max\{\lambda(c) \mid c \in A^*\} = \lambda(c') = d.$$

Proposition 5.5. *Let C be a complete biprefix code of degree d. If c is a quasi-power of order n, then we have*

$$\lambda(c) \geqslant \inf(n, d) \tag{1}$$

Proof. By induction over n. If $n = 0$, (1) holds.

Let us suppose that (1) holds for $0 \leqslant q < n$ and let c be a quasi-power of order n. Then, $c = xyx$, where x is a quasi-power of order $n-1$. By induction hypothesis we have

$$\lambda(x) \geqslant \inf(d, n-1).$$

This inequality and Proposition 5.4(i) imply

$$\lambda(c) = \lambda(xyx) \geqslant \lambda(x) \geqslant \inf(d, n-1). \tag{2}$$

If $\inf(d, n-1) = d$, then (1) holds. Suppose that $\inf(d, n-1) = n-1 < d$. If one of the inequalities in (2) is not an equality, we have $\lambda(c) \geqslant n$ and (1) holds. Finally, we show that the case

$$\lambda(c) = \lambda(xyx) = \lambda(x) = n-1 < d \tag{3}$$

cannot happen (and this completes the proof).

By contradiction, suppose that (3) holds. We prove by induction that (3) implies

$$\forall m \in \mathbb{N}: \quad \lambda((xy)^m x) = \lambda(x) = n-1 < d. \tag{4}$$

By hypothesis, (4) holds if $m = 1$. Suppose that (4) holds for $1 \leqslant r < m$. Then, by Proposition 5.3, we have

$$t \text{ prefix of } (xy)^r x,\ t \in S \Rightarrow |t| \leqslant |x|. \tag{5}$$

Now we shall prove that (5) is true also for $r = m$; that is,

$$t \text{ prefix of } (xy)^m x,\ t \in S \Rightarrow |t| \leqslant |x|. \tag{6}$$

Let t be a prefix of $(xy)^m x$. If t is a prefix of $(xy)^{m-1}x$, then, by (5) with $r = m-1$, (6) holds. Otherwise, $|t| > |(xy)^{m-1}x|$ and we prove that t does not belong to S.

Let us suppose, by contradiction, $t \in S$. Therefore, by hypothesis on $|t|$, there exists a $p \in A^+$, prefix of yx, such that $t = (xy)^{m-1}xp$. Set $v = (xy)^{m-2}xp$. Then, because v is a suffix of t, $v \in S$. However, since $p \in A^+$, v is a prefix of $(xy)^{m-1}x$ such that $|v| > |x|$. This contradicts (5) in the case of $r = m-1$. Then (6) holds and, by Proposition 5.3, (4) holds.

Thus, by Proposition 5.4(ii), $(xy)^m x \in H(C)$ for any $m \in \mathbb{N}$, and this contradicts the finiteness of C. □

Theorem 5.6. *There exists only a finite number of finite complete biprefix codes of a given degree d.*

Proof. Let C be a complete biprefix code of degree d. By Proposition 5.1, there exists a $k(d)$ such that any path c of length at least $k(d)$ has a quasi-power of order d as a factor, i.e., $c = c_1 y c_2$, where y is a quasi-power of order d.

Then, by Propositions 5.4(i) and 5.5, we have

$$\lambda(c) \geqslant \lambda(y) \geqslant \inf(d, d) = d.$$

Therefore, $\lambda(c) = d$. By Proposition 5.4(ii), c is not a factor of any element of C, i.e.,

$$\forall c \in A^*: \quad |c| \geqslant k(d) \Rightarrow c \notin H(C).$$

Then the length of the paths of a finite complete biprefix code of degree d is bounded by $k(d)+1$ and this implies the finiteness of this set of codes. □

6. The Césari–Schützenberger algorithm

In this section we extend an algorithm to the codes of paths that allows us to construct all the finite complete biprefix codes (Theorems 6.5 and 6.8). For this we need a notion of completion relative to a vertex.

Let v be a vertex. We say that a path starts from (ends in) v if it has v as initial (terminal) vertex.

Let C be a prefix (suffix) code. C is *v-complete* if the set of paths of C starting from (ending in) v is maximal among the prefix (suffix) codes of paths starting from (ending in) v. For example, on the graph from Fig. 1, the sets $\{c, d\}$ and $\{c, dd, dc\}$ ($\{a, c\}, \{a, aa, c\}$) are t_b-complete prefix (i_b-complete suffix) codes.

Let C be a complete biprefix code. For any $x \in A^*$ let us denote

$$Cx^{-1} = \{z \in A^* \mid zx \in C\}, \qquad x^{-1}C = \{z \in A^* \mid xz \in C\}.$$

A *good path* for C is a path x such that

(i) $\quad A^+x \cap C \neq \emptyset, \qquad xA^+ \cap C \neq \emptyset,$

(ii) $\quad (Cx^{-1})x \cap x(x^{-1}C) = \emptyset.$

For instance, if $A^2 = \{aa, ab, bd, bc, dd, dc, ca, cb\}$ is the homogeneous code on the graph from Fig. 1, we have that b is a good path for A^2.

Let C be a complete biprefix code, x a good path for C. The *transformed* $C(x)$ of C with respect to x is defined by

$$C(x) = C \cup x \cup (Cx^{-1})x(x^{-1}C) \setminus ((Cx^{-1})x \cup x(x^{-1}C)).$$

We shall also say that $C(x)$ is obtained from C by *internal transformation* with respect to the path x.

As in the case of the codes of words, by a similar argument [3], we have the following lemma.

Lemma 6.1. *Let X, Y be two v-complete prefix (suffix) codes of paths starting from v. Then X is different from Y if and only if a path of one code is a proper prefix (suffix) of a path of the other.*

In the following we shall use the next result.

Lemma 6.2. *Let C be a complete prefix (suffix) code, x a proper prefix (suffix) of an element of C. Then $x^{-1}C$ (Cx^{-1}) is a t_x-complete prefix (an i_x-complete suffix) code of paths starting from t_x (ending in i_x).*

Proof. The proof that $x^{-1}C$ is a prefix code is the same as for the codes of words. If $x^{-1}C$ is not t_x-complete, then a path $c \notin x^{-1}C$ exists such that $i_c = t_x$ and such that $x^{-1}C \cup c$ is a prefix code. Then $C \cup xc$ is a prefix code. Therefore, since $xc \notin C$, C is not a complete prefix code, contradicting the hypothesis.

In a symmetric way we prove the result for suffix codes. □

Let C be a complete biprefix code. A path $c \in C$ is an *internal path* if it is a proper factor of a path of C. Let us denote $\mathring{C}$ the *kernel* of C, i.e., the set of the internal paths of C:

$$\mathring{C} = \{c \in C \mid C \cap A^+ cA^+ \neq \emptyset\}.$$

We then have the following lemma.

Lemma 6.3. *Let C be a complete biprefix code. If C has no internal paths, then C is homogeneous.*

Proof. Suppose that $\mathring{C} = \emptyset$ and let n be the minimal length of the paths of C.

(1) Let us prove, by induction on $k \in \mathbb{N}$, that, for any $x \in A^*$ of length $n+k$, if the suffix of x of length n belongs to C, then the prefix of x of length n belongs to C. The statement holds for $k = 0$. Suppose that it holds for $0 \leqslant k' < k$. Let $a_1, \ldots, a_{n+k} \in A$, $x = a_1 \ldots a_{n+k} \in A^*$ and suppose that $y = a_{k+1} \ldots a_{k+n} \in C$: we must prove that $a_1 \ldots a_n \in C$. First we note that $z = a_k y$ cannot be a prefix of an element of C (this derives from $y \in C$, C suffix and $\mathring{C} = \emptyset$). Then, C being complete on the left, z has a proper prefix in C. By definition of n, this is $a_k \ldots a_{k+n-1}$. By induction hypothesis applied to $x' = a_1 \ldots a_{k+n-1}$, we have $a_1 \ldots a_n \in C$.

(2) Let $c \in A^n$. Since G is strongly connected, for any $c' \in C$ there is a $t \in A^*$ such that $ctc' \in A^*$. Let us suppose $|c'| = n$. By (1), we have $c \in C$. Therefore, $A^n \subseteq C$ and maximality of C implies $C = A^n$. □

Lemma 6.4. *Let $D \subseteq A^*$ be a prefix code, $B \subseteq A^*$ a t_d-complete prefix code for all $d \in D$. Let C_1, C_2 be two subsets of A^* such that $D \subseteq C_1$, $DB \subseteq C_2$ and $C_1 \setminus D = C_2 \setminus DB$. Then,*

(1) *C_1 is prefix if and only if C_2 is prefix;*

(2) *If C_1 is prefix, then C_1 is complete if and only if C_2 is a complete prefix code.*

Proof. (1): For any $d \in D$, dB is nonempty. In fact, let $d \in D$. Since G is strongly connected, there exists an arrow x starting from t_d. Since B is a t_d-complete prefix code, x is a prefix of an element $b \in B$ and $db \in DB$. Moreover, DB is a prefix code because otherwise there exist $d_1, d_2 \in D$, $b_1, b_2 \in B$ and $t \in A^+$ such that $d_1 b_1 t = d_2 b_2$. Then, since D is a prefix code, we have $d_1 = d_2$, and b_1 would be a proper prefix of b_2. This is a contradiction because B is a prefix code. Set $F = C_1 \setminus D = C_2 \setminus DB$. If F is not a prefix code, then neither C_1 nor C_2 are prefix codes. Therefore, suppose that F is a prefix code. Then we have

$$C_1 = F \cup D \quad \text{and} \quad C_2 = F \cup DB.$$

If C_1 is not a prefix code, then either a path f of F is a proper prefix of a path $d \in D$, or a path d of D is a proper prefix of a path f of F.

In the first case, since dB is not empty, f is a proper prefix of a path db of DB. Then C_2 is not a prefix code.

In the second case, let $dm = f$; m is not a path of B because F and DB are disjoint. Since $i_m = t_d$ and B is t_d-complete, we have either m a proper prefix of a path of B, or a proper prefix of m belonging to B. In both of cases, C_2 is not a prefix code.

Conversely, if C_2 is not a prefix code, there exists either a path $db \in DB$ proper prefix of a path $f \in F$, or a path $f \in F$ proper prefix of a path db of DB ($d \in D$, $b \in B$). In the first case, d is a proper prefix of f and C_1 is not a prefix code. In the second case, d is a proper prefix of f or f is a proper prefix of d (since D and F are disjoint); in both of cases, C_1 is not a prefix code.

(2): By (1), C_1 is prefix if and only if C_2 is prefix. If C_1 is not complete, there exists a $z \in A^+ \setminus C_1$ such that $C_1 \cup z$ is a prefix code. Now z is not a path of C_2 and $C_2 \cup z$ is a prefix code (by (1) applied to $C_1' = C_1 \cup z$ and $C_2' = C_2 \cup z$). Then C_2 is not complete.

Conversely, if C_2 is not complete, there exists a path z of $A^+ \setminus C_2$ such that $C_2 \cup z$ is a prefix code. Now z is not a path of C_1 and $C_1 \cup z$ is a prefix code (by (1) applied to $C_1' = C_1 \cup z$ and $C_2' = C_2 \cup z$). Therefore, C_1 is not complete. By Propositions 5.3, 5.4(ii) and Lemma 6.2, we deduce that $C(x)$ has the same degree as C. □

Theorem 6.5. *The transformed $C(x)$ of a complete biprefix code C by a good path*:

$$C(x) = C \cup x \cup (Cx^{-1})x(x^{-1}C) \setminus ((Cx^{-1})x \cup x(x^{-1}C))$$

is a complete biprefix code having the same degree as C.

Proof. By Lemma 6.4 applied to $C_1 = C \cup x \setminus x(x^{-1}C)$ and $C_2 = C$ (with $B = x^{-1}C$, $D = x$), we have that C_1 is a complete prefix code. By the same lemma applied to $C_1 = C \cup x \setminus x(x^{-1}C)$ and to $C_2 = C(x)$ (with $B = (x^{-1}C)$, $D = (Cx^{-1})x$), we have that $C(x)$ is a complete prefix code. The symmetry of the construction allows the conclusion. By Propositions 5.3, 5.4(ii) and Lemma 6.2, we deduce that $C(x)$ has the same degree as C. □

The following lemma is crucial in proving that any biprefix code can be obtained from a homogeneous code by a finite number of internal transformations (see [3, Theorem 1]).

Lemma 6.6. *Let C be a complete biprefix nonhomogeneous code. There exist an $x \in \mathring{C}$, a t_x-complete prefix code R of paths starting from t_x, an i_x-complete suffix code T of paths ending in i_x such that $TxR \subseteq C$.*

Proof. By Lemma 6.3, there exists an $x \in \mathring{C}$ of maximal length. Therefore, the set

$$\{u \in A^+ \mid (ux)^{-1}C \neq \emptyset\}$$

is not empty.

(1) We note that the statement is true if we suppose that there exists a $u \in A^+$ with $(ux)^{-1}C \neq \emptyset$ and such that

$$\forall p, y \in (ux)^{-1}C: \quad C(xp)^{-1} = C(xy)^{-1}. \tag{7}$$

In fact, take $p \in (ux)^{-1}C$ and set

$$R = (ux)^{-1}C, \qquad T = C(xp)^{-1}.$$

Let $t \in T$, $r \in R$. Since $r \in R = (ux)^{-1}C$, by (7) we have that $T = C(xp)^{-1} = C(xr)^{-1}$. Then $t \in T$ belongs to $C(xr)^{-1}$, that is, $txr \in C$. This proves that $TxR \subseteq C$. Moreover, by Lemma 6.2, R is a t_x-complete prefix code and T is an i_x-complete suffix code.

(2) By contradiction, suppose that the statement is not true. Then, by (1), for any $u \in A^+$ such that $(ux)^{-1}C \neq \emptyset$ we have

$$\exists p, y \in (ux)^{-1}C: \quad C(xp)^{-1} \neq C(xy)^{-1}$$

and choose the triple (u, p, y) such that $|p|+|y|$ is minimal.

By Lemma 6.1, there exists a $z \in C(xp)^{-1} \setminus C(xy)^{-1}$ which is a proper suffix of an element of $C(xy)^{-1}$, i.e.,

$$\exists z' \in A^+: \quad z'z \in C(xy)^{-1}. \tag{8}$$

By (8), $z'zxy \in C$ and, by Lemma 6.2, $(z'zx)^{-1}C$ is a t_x-complete prefix code. Since $i_p = t_x$, either a prefix of p is in $(z'zx)^{-1}C$ or p is a prefix of an element of $(z'zx)^{-1}C$. We prove that this contradicts the hypothesis on C, x, $|p|+|y|$ and completes the proof.

In fact, since $z \in C(xp)^{-1}$, and $C(xp)^{-1}$ is a suffix code, we have

$$z'z \notin C(xp)^{-1} \Rightarrow z'zxp \notin C \Rightarrow p \notin (z'zx)^{-1}C.$$

Moreover, p cannot be a proper prefix of an element of $(z'zx)^{-1}C$ because otherwise there exists a $c \in A^+$ such that

$$pc \in (z'zx)^{-1}C \Rightarrow z'zxpc \in C.$$

Since z', $c \in A^+$ and $z \in C(xp)^{-1}$, we have $zxp \in \mathring{C}$ which contradicts maximality of $|x|$.

Finally, suppose that a proper prefix of p is in $(z'zx)^{-1}C$, i.e.,

$$\exists c_1 \in (z'zx)^{-1}C,\ c_2 \in A^+: \quad p = c_1c_2. \tag{9}$$

Then,

$$y \in (ux)^{-1}C \Rightarrow u \in C(xy)^{-1}. \tag{10}$$

Since $(ux)^{-1}C$ is a prefix code, we have

$$p \in (ux)^{-1}C \Rightarrow c_1 \notin (ux)^{-1}C \Rightarrow uxc_1 \notin C \Rightarrow u \notin C(xc_1)^{-1}. \tag{11}$$

Then, by (8) and (9), we have that $y, c_1 \in (z'zx)^{-1}C$ and, by (10) and (11), $C(xy)^{-1} \neq C(xc_1)^{-1}$. Since $|y|+|c_1|<|y|+|p|$, the triple $(z'z, y, c_1)$ contradicts the minimality of $|y|+|p|$. □

Lemma 6.7. *Let C be a complete biprefix nonhomogeneous code. Then there exist $x \in \mathring{C}$, a t_x-complete prefix code R of paths starting from t_x and an i_x-complete suffix code T of paths ending in i_x such that*

(1) $Tx \cap xR = \emptyset$;

(2) $F = C \cup Tx \cup xR \setminus (x \cup TxR)$ *is a complete biprefix code*;

(3) x *is a good path for F and C is obtained from F by internal transformation with respect to x.*

Proof. By Lemma 6.6, there exist an $x \in \mathring{C}$, a t_x-complete prefix code R of paths starting from t_x and an i_x-complete suffix code T of paths ending in i_x such that

(*) $\quad TxR \subseteq C$

(1): If the sets Tx and xR are not disjoint, there exists a path $r \in R$ such that $xr \in Tx$. Then $t_r = t_x$ and, by (*), we have

$$xrR \subseteq TxR \subseteq C.$$

This is a contradiction, because $x \in C$ and C is biprefix.

(2): First we prove that Tx is a prefix subset of A^+. In fact, otherwise there would exist $t \in T$, $x_1 \in A^+$ such that $txx_1 \in Tx$. Then we have

$$txR \subseteq TxR \subseteq C, \qquad txx_1R \subseteq TxR \subseteq C$$

and, consequently,

$$R \subseteq (tx)^{-1}C, \qquad R \subseteq x_1^{-1}(tx)^{-1}C.$$

By hypothesis and Lemma 6.2, R, $(tx)^{-1}C$, $(x_1)^{-1}(tx)^{-1}C$ are t_x-complete codes of paths starting from t_x. Then we have

$$R = (tx)^{-1}C = x_1^{-1}(tx)^{-1}C.$$

These equalities and $t_{x_1} = t_x = i_r$ (for any $r \in R$) imply $x_1R = R$, contradicting the hypothesis of finiteness of R. Then Tx is a prefix subset of A^+. Moreover, by Lemma 6.4 applied to $C_1 = C$, $C_2 = (C \cup xR)\setminus x$ (with $D = x$, $B = R$), C_2 is a complete prefix code. By the same lemma applied to $C_1 = F$ and to $C_2 = (C \cup xR)\setminus x$ (with $D = Tx$, $B = R$), F is a complete prefix code.

By symmetry of the construction, F is a complete biprefix code.

(3): Since $xR \subseteq F$, we have $R \subseteq x^{-1}F$. But, by Lemma 6.2, R and $x^{-1}F$ are two t_x-complete prefix codes of paths starting from t_x. Then $R = x^{-1}F$. By a symmetrical

argument, we have $T = Fx^{-1}$. Then, by (1) and (2), x is a good path for F and, by (2), C is obtained from F by internal transformation with respect to x. □

Theorem 6.8. *Any finite complete biprefix code C of degree d can be obtained by a finite number of internal transformation starting with the homogeneous code A^d.*

Proof. Let $C_0 = C$ and, for any $h \in \mathbb{N}$, $h > 0$, if C_{h-1} is not a homogeneous code, denote by C_h the code obtained by Lemma 6.7(2) applied to $C = C_{h-1}$:

$$C_h = C_{h-1} \cup T_{h-1}x_{h-1} \cup x_{h-1}R_{h-1} \setminus (x_{h-1} \cup T_{h-1}x_{h-1}R_{h-1})$$

(where $x = x_{h-1}$, $R = R_{h-1}$, $T = T_{h-1}$ are defined as in Lemma 6.7).

(1) By induction we can see that, for any $h \in \mathbb{N}$, the elements of C_h are factors of $C = C_0$. Then, since C is finite, we have a finite number of codes C_h.

(2) If we denote by $\ell(C_h)$ the sum of the lengths of the paths of C_h, it is clear that, for any $h \in \mathbb{N}$, we have

$$\ell(C_{h+1}) \leqslant \ell(C_h),$$

$$|R_h| > 1 \text{ or } |T_h| > 1 \Rightarrow \ell(C_{h+1}) < \ell(C_h).$$

(3) In order to prove the statement we must show that there exist k, $t \in \mathbb{N}$ such that $C_k = A^t$. By contradiction, suppose that this is not true. Then, by (1), there exist i, $n \in \mathbb{N}$, $n \geqslant 2$ such that $C_i = C_{i+n}$. Therefore, by (2), for any $j \in \{0, \ldots, n-1\}$, we have

$$|R_{i+j}| = |T_{i+j}| = 1.$$

Denote by r_{i+j}, t_{i+j} paths such that $R_{i+j} = \{r_{i+j}\}$, $T_{i+j} = \{t_{i+j}\}$ and E_j is the $\mathbb{N}$-set of lengths of the paths of C_{i+j} (we suppose that any length appears in E_j with the same multiplicity as in C_{i+j}). By

$$C_{i+j+1} = C_{i+j} \cup t_{i+j}x_{i+j} \cup x_{i+j}r_{i+j} \setminus (x_{i+j} \cup t_{i+j}x_{i+j}r_{i+j}),$$

we have

$$E_{j+1} = E_j \cup |t_{i+j}x_{i+j}| \cup |x_{i+j}r_{i+j}| \setminus (|x_{i+j}| \cup |t_{i+j}x_{i+j}r_{i+j}|).$$

Let us define $|x_{i+j}|$, $|t_{i+j}x_{i+j}r_{i+j}|$ the *active elements* and let ℓ be the maximum in the set of the active elements, and m_j the multiplicity of ℓ in E_j.

(a) For all $j \in \{0, \ldots, n-1\}$ we have $m_j \geqslant m_{j+1}$ (otherwise there exists a $j \in \{0, \ldots, n-1\}$ such that $m_j < m_{j+1}$; then, by definition of E_{j+1}, either $\ell = |t_{i+j}x_{i+j}|$ or $\ell = |x_{i+j}r_{i+j}|$ and $\ell' = |t_{i+j}x_{i+j}r_{i+j}|$ is an active element greater than ℓ).

(b) There exists a $q \in \{0, \ldots, n-1\}$ such that $m_q > m_{q+1}$ (in fact, since ℓ is active, there is a $q \in \{0, \ldots, n-1\}$ such that $\ell = |t_{i+q}x_{i+q}r_{i+q}|$ and the preceding inequality follows by definition of E_{q+1}).

(c) By (a) and (b), we have

$$m_0 \geqslant m_1 \geqslant \cdots \geqslant m_q > m_{q+1} \geqslant \cdots \geqslant m_n.$$

Then, $m_0 \neq m_n$. On the other hand,

$$C_i = C_{i+n} \Rightarrow E_0 = E_n \Rightarrow m_0 = m_n.$$

This contradiction completes the proof. □

For example, the code $C = \{aa, abc, b, abd, dc, ca, cbc, dd, cbd\}$ on the graph from Fig. 1 can be obtained from the homogeneous code $A^2 = \{aa, ab, bd, dc, dd, bc, cb, ca\}$ by internal transformation with respect to the path b:

$$A^2(b) = A^2 \cup b \cup (A^2b^{-1})b(b^{-1}A^2)\setminus((A^2b^{-1})b \cup b(b^{-1}A^2)) = C.$$

The code $D = \{aa, ab, bc, bd, ca, cb, e, ded, dec\}$ on the graph from Fig. 4 can be obtained from the homogeneous code $B^2 = \{aa, ab, bc, bd, ca, cb, de, ed, ec\}$ by internal transformation with respect to the path e:

$$B^2(e) = B^2 \cup e \cup (B^2e^{-1})e(e^{-1}B^2)\setminus((B^2e^{-1})e \cup e(e^{-1}B^2)) = D.$$

7. Derivation

In this section we show that we can associate to any code C of degree d its *derivative*, a code C' of degree $d-1$ (Theorem 7.3). This construction extends the derivation of codes of words as defined by Césari [4] to the case of paths.

Let C be a finite complete biprefix code of degree $d \geq 2$, $\mathring{C}$ the *kernel* of C, $H = H(C)$ the set of proper factors of C, and P (S) the set of proper prefixes (suffixes) of the elements of C. The *derivative* C' of C is defined as follows:

$$C' = \mathring{C} \cup (P \cap S\setminus H).$$

For example, if C is the code

$$\{aa, abc, b, abd, dc, ca, cbc, dd, cbd\}$$

on the graph from Fig. 1 we have $\mathring{C} = \{b\}$ and $C' = A = \{a, b, c, d\}$ (the homogeneous code of degree 1). Let D be the code $\{aa, ab, bc, bd, ca, cb, e, ded, dec\}$ over the graph from Fig. 4; one can see that $\mathring{D} = \{e\}$ and $C' = B = \{a, b, c, d, e\}$.

Remark 7.1. (a) We stress the existence of the following inclusion:

$$H(C') \subseteq H(C).$$

In fact, if $c \in H(C')$, then there exist $u, p \in A^+$ such that $ucp \in C'$. Then either $ucp \in \mathring{C}$ or $ucp \in P \cap S\setminus H$. In both cases, we have $c \in H(C)$.

(b) For any subset M of A^*, let us denote by 1_M the set of the empty paths in M. We stress the existence of the following inclusion:

$$1_P = 1_S \subseteq 1_H. \qquad (*)$$

Indeed, let s be an empty path in S. Since G is strongly connected, there exists $t' \in A$ such that $st' \in A^*$. Since C is a complete prefix code, st' is a prefix of an element of C. As a consequence $s \in 1_P$. Moreover, we have two cases:

$$\exists t \in A: \quad ts \in S,$$

$$\exists t \in A: \quad ts \in C.$$

Since C is biprefix and complete, in the first case ts is a proper prefix of an element of C. Consequently $s \in 1_H$. Suppose that the second case holds. Since G is strongly connected for any $x \in A$ there exists $z_x \in A^*$ such that $tsz_x x \in A^*$. Since C is a complete suffix code, for any $x \in A$ and for any nonempty prefix z of $z_x x$, either tsz has a suffix in C or tsz is a suffix of an element of C. Then we have two cases:

(1) $\exists x \in A\ \exists z$, z nonempty prefix of $z_x x$: tsz is a (proper) suffix of an element of C,

(2) $\forall x \in A$, for any nonempty prefix z of $z_x x$, z has a suffix in C.

Since C is prefix, in the second case we have $x \in C$, for any $x \in A$, i.e. $C = A$. This is a contradiction since the degree of C is at least two. In the first case we have $s \in 1_H$. In a symmetrical way if $p \in 1_P$ is an empty path we have $p \in 1_S \cap 1_H$. By $(*)$ we have $C' \leqslant A^+$.

Let S' be the set of proper suffixes of C'. We then have the following lemma.

Lemma 7.2. *If C is a finite complete biprefix code, then we have $S' = S \cap H$.*

Proof. We prove that $S' \subseteq S \cap H$. Let $s \in S'$. Then there exists a $t \in A^+$ such that either $ts \in \mathring{C}$ or $ts \in P \cap S \setminus H$.

In the first case, there exist $u, z \in A^+$ such that ts and $utsz$ belong to C. Then $s \in S \cap H$.

In the second case, since $ts \in S$, there exists a $u \in A^+$ such that

$$uts \in C \tag{12}$$

and since $ts \in P$, there exists a $z \in A^+$ such that

$$tsz \in C. \tag{13}$$

By (12), we have $s \in S$ and, by (13), we have $s \in H$.

Conversely, let us prove that $S \cap H \subseteq S'$. Let $s \in S \cap H$. Then, since $s \in H$, there exist a t of maximal length and a z in A^+ such that

$$tsz \in C. \tag{14}$$

Moreover, since $s \in S$, by Lemma 6.2, Cs^{-1} is a suffix i_s-complete code. Hence, since $t \notin Cs^{-1}$ (C is a biprefix code) and since $t_t = i_s$, there exists a $u \in Cs^{-1}$, i.e.,

$$us \in C \tag{15}$$

such that either u is a proper suffix of t or t is a proper suffix of u.

In the first case, by (14) and (15) we have $us \in \mathring{C} \subseteq C'$ and then $s \in S'$.

In the second case, let us prove that we have $ts \in C'$. In fact, in this case, ts is a proper suffix of us and, by (15), $ts \in S$. On the other hand, by (14), $ts \in P$. Finally, $ts \notin H(C)$ because $ts \in H(C)$ contradicts the maximality of $|t|$. Then $ts \in P \cap S \setminus H \subseteq C'$ which implies $s \in S'$. □

Theorem 7.3. *The derivative C' of a finite complete biprefix code C of degree d is a finite complete biprefix code of degree $d-1$.*

Proof. Let us prove that C' is a biprefix code. By contradiction, let $x, y \in C'$, $t \in A^+$ such that $xt = y$. Then we have the following four cases:

(a) $x, y \in \mathring{C}$;

(b) $x \in \mathring{C}, y \in P \cap S \setminus H$;

(c) $x \in (P \cap S) \setminus H, y \in \mathring{C}$;

(d) $x, y \in (P \cap S) \setminus H$.

We cannot have case (a) because $\mathring{C} \subseteq C$ and C is a prefix code.

Case (b) cannot hold because, since $y \in P$, there exists a $z \in A^+$ such that $xtz = yz \in C$. Then $x \in \mathring{C} \subseteq C$ should be a proper prefix of $yz \in C$, but this is a contradiction because C is a prefix code.

We cannot have case (c) because since $y \in \mathring{C}$, there exist $u, z \in A^+$ such that $uxtz = uyz \in C$, which implies $x \in H(C)$, a contradiction.

Finally, case (d) cannot hold because since $y \in S$, there exists a $u \in A^+$ such that $uxt = uy \in C$, which implies $x \in H(C)$ and this is a contradiction.

Then C' is a prefix code and, by symmetry of the construction, a biprefix code. Let us prove that C' is a complete code.

(1) First if $c \in C$, then either $c \in \mathring{C}$ or c has a prefix in C'. In fact, if $c \notin \mathring{C}$, let s be the prefix of c of maximal length belonging to S (it exists since $1 \in S$). Then $s \neq c$ (because C is biprefix) and there exists an $m \in A^+$ such that

$$c = sm. \tag{16}$$

Let us prove that $s \in C'$. We have $s \in P \cap S$. Moreover, $s \notin H(C)$. In fact, otherwise there should exist $u, z \in A^+$ such that $usz \in C$. By Lemma 6.2, $(us)^{-1}C$ is a t_s-complete prefix code. Then, since $m \notin (us)^{-1}C$ (C is biprefix), either m has a proper prefix in $(us)^{-1}C$ or it is a proper prefix of an element of $(us)^{-1}C$. By (16), the first case contradicts maximality of $|s|$, and the second case contradicts $c \notin \mathring{C}$.

(2) Let $x \in A^+$. Since C is complete, there exist $c \in C$, $t \in A^+$ such that either $x = ct$ or $xt = c$. In both of the cases, by (1), either x has a prefix in C' or it is a prefix of an element of C'. This proves that C' is a complete prefix code. By symmetry of the construction we have that C' is a complete suffix code.

Finally, we show that if the degree of C is d, then the degree of C' is $d-1$. Let $c \in A^*$, $c \notin H(C)$. Then, by Remark 7.1, we have $c \notin H(C')$. By Proposition 5.4(ii) we must prove that c has exactly $d-1$ C'-interpretations, i.e., by Proposition 5.3, that the number of the prefixes of c belonging to S' is $d-1$. Since $c \notin H(C)$, by

Proposition 5.4(ii), c has d C-interpretations. Moreover, by Proposition 5.3, d is the number of the prefixes $s_1 < s_2 < \cdots < s_d$ of c belonging to S (where we denote by $<$ the relation "to be a proper prefix of"). By Lemma 7.2, we have $S' \subseteq S$ and the set of the prefixes of c belonging to S' is a subset of $\{s_1, \ldots, s_d\}$. Let us prove that $s_1, \ldots, s_{d-1} \in S'$ and $s_d \notin S'$. In fact, since $s_d \in S$, there exists a $t \in A^+$ such that $ts_d \in C$. Therefore, for any $i \in \{1, \ldots, d-1\}$, s_i is a proper factor of $ts_d \in C$ and, by Lemma 7.2, we have

$$\forall i \in \{1, \ldots, d-1\}: \quad s_i \in H(C) \cap S = S'.$$

Moreover, s_d does not belong to $H(C)$ (and, consequently, $s_d \notin S'$). In fact, let us suppose by contradiction that $s_d \in H(C)$. Then there exist $u, z \in A^+$ such that $us_dz \in C$. Then $s_1, s_2, \ldots, s_d, s_dz$ are $d+1$ prefixes of s_dz belonging to S. This contradicts the hypothesis that C has degree d. □

8. Integration

In this section we are going to prove that, given a finite complete biprefix code C of degree d, we can construct a finite complete biprefix code of degree $d+1$ whose derivative is C (Theorem 8.12). Thus any finite complete biprefix code can be "integrated" as in the case of the codes of words [4]. First, we give some definitions and lemmas. As in the case of codes of words, by an analogous argument [4], we have the following proposition.

Proposition 8.1. *Let C be a complete suffix code. For any $c \in C$, $D = (C \cup Ac) \setminus c$ is a complete suffix code.*

A biprefix subset X is *sufficient* if there exists a double-infinite path z such that for any point of z there is a decoding of z in X passing through it. In this case, we also say that X *gives all decodings* of z. Then a biprefix subset X is *insufficient* if, for any double-infinite path, X does not give all its decodings. A path c is *full* (with respect to X) if there is an X-interpretation passing through any point of c.

Remark 8.2. The set T of the full paths with respect to a biprefix subset is a factor-closed set (i.e., if c is a full path, then any factor of c is full). Then *if T is infinite, there is an infinite sequence $L = (\ell_n)_{n \in \mathbb{N}}$ of paths of T such that, for any $k \in \mathbb{N}$, ℓ_k is a proper factor of ℓ_{k+1}.*

By induction, for any $k \in \mathbb{N}$ there exists an $\ell_k \in T$ such that the set L_k of the paths of T having ℓ_k as a proper factor is infinite and, for any $k > 0$, ℓ_{k-1} is a proper factor of ℓ_k.

Since T is infinite there is an $\ell_0 \in A$ such that the set L_0 of the paths of T having ℓ_0 as a proper factor is infinite. Then $\ell_0 \in T$. Since L_0 is infinite, there exist $t_0, q_0 \in A$ such that the set L_1 of paths of T having $t_0\ell_0q_0$ as a proper factor is infinite. By setting $\ell_1 = t_0\ell_0q_0$ we have $\ell_1 \in T$ and the statement holds for $k = 0$ and $k = 1$.

Suppose that the statement is true for $k \geqslant 1$. Since L_k is infinite, there exists $t_k, q_k \in A$ such that the set L_{k+1} of the paths of T having $t_k \ell_k q_k$ as a proper factor is infinite. By setting $\ell_{k+1} = t_k \ell_k q_k$ we have $\ell_{k+1} \in T$ and the statement holds for $k+1$.

Consider A provided with the discrete topology and the set ${}^{\omega}A^{\omega}$ of the double-infinite paths provided with the product topology. ${}^{\omega}A^{\omega}$ is a metrizable compact space. Then each sequence of double-infinite paths has a subsequence that converges in ${}^{\omega}A^{\omega}$.

Lemma 8.3. *A biprefix set X of A^* is sufficient if and only if the set of the full paths with respect to X is infinite.*

Proof. If X is sufficient, then there is a double-infinite path z whose X gives all the decodings. Then any factor of z is a full path with respect to X, and this set is infinite.

Conversely, suppose that the set of the full paths with respect to X is infinite. By Remark 8.2, there is an infinite sequence $L = (\ell_n)_{n \in \mathbb{N}}$ of full paths such that, for any $n \in \mathbb{N}$, ℓ_n is a proper factor of ℓ_{n+1}. For any $n \in \mathbb{N}$, let t_n be a double-infinite path such that

$$\exists h_n, k_n \in \mathbb{Z},\ h_n < 0,\ k_n \geqslant 0: \quad t_n[h_n, k_n] = \ell_n,$$

$$\forall i < n: \quad t_n[h_i, k_i] = \ell_i.$$

By definition of $(\ell_n)_{n \in \mathbb{N}}$, for any $n \in \mathbb{N}$ we have $k_n < k_{n+1}$ and $h_n > h_{n+1}$. The sequence $(t_n)_{n \in \mathbb{N}}$ has a subsequence that converges in ${}^{\omega}A^{\omega}$ to ℓ. It is clear that ℓ is the double-infinite path obtained by "bottling" the elements of L, i.e.,

$$\forall n \in \mathbb{N}: \quad \ell[h_n, k_n] = \ell_n.$$

We claim that X gives all decodings of ℓ.

Let $j \geqslant 0$ be a point of ℓ and η the maximal length of paths in X. Since $(k_n)_{n \in \mathbb{N}}$ ($(h_n)_{n \in \mathbb{N}}$) is strictly increasing (decreasing) there exists an $m \in \mathbb{N}$ such that

$$k_m > j + \eta, \qquad h_m < -\eta.$$

Then $(z[h_m, j-1], z[j, k_m])$ is a point of ℓ_m, a full path with respect to X. Therefore, by definition of k_m, h_m, there exist an element of X, prefix of $z[j, k_m]$ and an element of X, suffix of $z[h_m, j-1]$.

By a similar argument, if $j < 0$, there exists an $m \in \mathbb{N}$ such that an element of X is prefix of $z[j, k_m]$ and an element of X is suffix of $z[h_m, j-1]$.

Since X is biprefix, this implies that for any point of ℓ there is a decoding of ℓ passing through it in X. □

Lemma 8.4. *Let C be a finite complete biprefix code, $\mathring{C}$ the kernel of C (i.e., the set of the internal paths of C), and I an insufficient subset of A^* such that*

$$\mathring{C} \subseteq I \subseteq C. \tag{17}$$

Let $\{D_i\}$ *be the sequence of paths defined by*

$$D_1 = I \cup A(C \setminus I),$$

$$\forall i > 1: \quad D_i = \begin{cases} D_{i-1} & \text{if, for any } c \in D_{i-1},\ c \text{ has} \\ & \text{no proper prefix in } I, \\ (D_{i-1} \cup Ac) \setminus \{c\} & \text{if } c \text{ is an element of } D_{i-1} \text{ having} \\ & \text{a proper prefix in } I. \end{cases}$$

Then there exists an $i > 1$ *such that* $D_i = D_{i-1}$.

Proof. By contradiction, suppose that for any $i > 1$ we have $D_i \neq D_{i-1}$. Then there exist a sequence of paths $(a_n)_{n \in \mathbb{N}}$ and a $y \in C \setminus I$ such that for any $k \in \mathbb{N}$ the sequence $a_k a_{k-1} \dots a_1 y$ is a path having a proper prefix in I.

(1) Let $L = (\ell_k)_{k \in \mathbb{N}}$ be the sequence of paths defined as follows:

$$\forall k \geqslant 2: \quad \ell_k = a_k \dots a_2$$

and, for any $k \geqslant 2$, let (u, v) be a point of ℓ_k. By setting $a_h \dots a_m = 1$ if $h < m$, we can suppose that

$$\exists t \in \{0, \dots, k\}: \quad u = a_k \dots a_{t+1}, \qquad v = a_t \dots a_2.$$

Let us prove that either u is suffix of an element of I or u has a suffix in I, and that either v is prefix of an element of I or v has a prefix in I. Since vy has a prefix in I, either v has a prefix in I or it is prefix of an element of I.

Let r be an integer such that $u_1 = a_r \dots a_{k+1} u$ is a path of length greater than the maximal length of the paths in C. Since C is a complete biprefix code, u_1 has a suffix in C. Then there exists a $q \in \mathbb{N}$ such that

$$u_1 = a_r \dots a_q \dots a_{t+1} \quad \text{and} \quad a_q \dots a_{t+1} \in C.$$

Since $a_q \dots a_{t+1} \dots a_1 y$ has a prefix in $I \subseteq C$ and since C is biprefix, this prefix is $a_q \dots a_{t+1}$. Then $a_q \dots a_{t+1} \in I$ and either u has a suffix in I or u is suffix of an element of I.

(2) Since I is biprefix, by (1), we have that for any $\ell_k \in L$ and for any point of ℓ_k there is an I-interpretation passing through it, i.e., ℓ_k is full with respect to I.

By Lemma 8.3, since L is infinite, I should be sufficient, contradicting the hypothesis. This contradiction concludes the proof. □

Let us denote by D the set obtained in Lemma 8.4 by a finite complete biprefix code C of degree d and by an insufficient set I satisfying (17). By Proposition 8.1, D is a finite complete suffix code. We show that D is a finite complete biprefix code of degree $d+1$ having kernel I and C as derivative. First we prove the following lemma.

Lemma 8.5. *For any* $u, v \in A^+$ *and* $x \in C \setminus I$ *we have* $uxv \notin D$.

Proof. By contradiction. Suppose that there are $u, v \in A^+$, $x \in C \setminus I$ such that $uxv \in D$. By definition of D, we have one of the following cases:

(1) $uxv \in I$,

(2) $\exists y \in C \setminus I, a_1, a_2, \ldots, a_k \in A$: $uxv = a_k \ldots a_1 y$

with $a_1 y, a_2 a_1 y, \ldots, a_{k-1} \ldots a_1 y$ having a prefix in I.

In the first case, $uxv \in I \subseteq C$ and $x \in C$ imply $x \in \mathring{C}$. This is a contradiction since $\mathring{C} \subseteq I$ and $x \in C \setminus I$.

In the second case, we note that we cannot have $y = xv$ (since $x, y \in C$, $v \in A^+$ and C is biprefix).

Then we have two cases:

(i) $|y| < |xv|$,

(ii) $|y| > |xv|$.

In case (i), since $u \in A^+$, we have $k > 1$ and there is a j, $1 \leqslant j < k$, such that $xv = a_j \ldots a_1 y$. By definition, $a_j \ldots a_1 y$ has a prefix $z \in I \subseteq C$ and $z \neq x$ since $x \in C \setminus I$. This is absurd since C is a biprefix code.

In case (ii) there is a $w \in A^+$ such that $wxv = y \in C \setminus I$ with $x \in C$. Then this equation implies $x \in \mathring{C} \subseteq I$, contradicting the hypothesis $x \in C \setminus I$. □

Lemma 8.6. *Let c be a nonempty path such that*

(i) *c has no prefix in I and c is not a prefix of an element of I.*

If there exists a proper suffix z of c such that

(ii) *either z has a prefix in $C \setminus I$ or z is prefix of an element of $C \setminus I$,*

then,

(iii) *either c is a prefix of an element of D or c has a prefix in D.*

Proof. By hypothesis, there are $z \in A^*$, $u \in A^+$, $v \in A^*$, $x \in C \setminus I$ such that either

$$c = uxv, \qquad z = xv$$

or

$$cv = ux, \qquad zv = x$$

and let $|z|$ be maximum with respect to (ii). If we prove $ux \in D$, the statement holds.

By definition, $ux \in D$ if

(1) ux has no prefix in I;

(2) for any proper nonempty suffix u_1 of u, $u_1 x$ has a prefix in I.

By (i), (1) holds. Let u_1 be a proper nonempty suffix of u. Since C is a complete prefix code, we have $u_1 z A^* \cap C^* \neq \emptyset$. Then there are $x_1 \in C$, $v_1 \in A^*$ such that either $u_1 z = x_1 v_1$ or $u_1 z v_1 = x_1$. Since $|u_1 z| > |z|$, by hypothesis on $|z|$, we have $x_1 \in I$. Therefore, we have four cases:

(a) $u_1 z = x_1 v_1 = u_1 xv$

(b) $u_1 zv = x_1 v_1 v = u_1 x$

(c) $u_1 z v_1 = x_1 = u_1 x v v_1$

(d) $u_1 z v_1 = x_1$, $zv = x$.

In case (a), by Lemma 8.5, we have $|x_1| \leqslant |u_1 x|$ and (2) holds.

In case (b), x_1 is a prefix of $u_1 x$ and (2) holds.

Case (c) with $vv_1 \in A^+$ cannot happen because of Lemma 8.5. Then $vv_1 = 1$ and x_1 is prefix of u_1x, and thus (2) holds.

In case (d), $v_1 = 1$ implies x_1 prefix of u_1x (and (2) holds). On the other hand, if $v = 1$, by Lemma 8.5, $v_1 = 1$ and x_1 is a prefix of u_1x (and (2) holds).

Thus, we can suppose $v \neq 1$, $v_1 \neq 1$. We have $v \notin (u_1z)^{-1}C$ (otherwise, $u_1zv = u_1x \in C$ with $u_1 \in A^+$, $x \in C \setminus I \subseteq C$, a contradiction since C is biprefix). Moreover, by Lemma 6.2, since u_1z is a proper prefix of $x_1 \in I \subseteq C$, $(u_1z)^{-1}C$ is a t_z-complete prefix code. Since $t_z = i_v$, there exist $m \in A^+$, $y \in (u_1z)^{-1}C$ (i.e., $u_1zy \in C$) such that either $v = ym$ or $vm = y$. The case $vm = y$ cannot happen (indeed, $vm = y$ implies $u_1zvm = u_1zy \in C$ with $u_1, m \in A^+$ and $zv = x \in C \setminus I$. Then $x \in \mathring{C} \subseteq I$: a contradiction). Therefore, we have $v = ym$ which implies $u_1x = u_1zv = u_1zym$. Since $u_1zy \in C$ and $u_1zy \notin C \setminus I$ (otherwise, u_1z is a proper suffix of c, satisfying (ii) and such that $|u_1z| > |z|$, contradicting the definition of z), we have $u_1zy \in I$ is a prefix of u_1x and (2) holds. □

Lemma 8.7. *Let c be a nonempty path. There exist a $t \in A^*$ such that $ct \in A^*$ and a proper suffix z of ct satisfying*

(i) *either z has a prefix in $C \setminus I$ or z is a prefix of an element of $C \setminus I$.*

Proof. Let $x \in A$, $c' \in A^*$ such that $c = xc'$. Since G is strongly connected, there is a sequence $(t_n)_{n \in \mathbb{N}}$ of paths with $t_0 = 1$ and such that

(1) $\forall n \in \mathbb{N}$: t_n is a proper prefix of t_{n+1};

(2) $c't_n \in A^*$.

Suppose that the statement does not hold. Then,

(ii) For any $n \in \mathbb{N}$, any suffix of $c't_n$ is not a prefix of an element of $C \setminus I$ and it has no prefix in $C \setminus I$.

Since $(|t_n|)_{n \in \mathbb{N}}$ is strictly increasing, let $m \in \mathbb{N}$ be such that $\forall n \geqslant m$: $|c't_n|$ is greater than the maximal length $\ell(C \setminus I)$ of the paths in $C \setminus I$, i.e.,

$$\forall n \geqslant m, \exists t'_n \in A^+, p \in A^*, |p| \geqslant \ell(C \setminus I): \quad c't_n = pt'_n.$$

Let (y_1, y_2) be a point of t'_n. Then (py_1, y_2) is a point of pt'_n. By Proposition 5.2, we have

(a) either y_2 has a prefix in C or y_2 is a prefix of an element of C;

(b) either py_1 is suffix of an element of C or py_1 has a suffix in C.

Since y_2 is a suffix of $c't_n$ by (ii) and (a), we have

(a') either y_2 has a prefix in I or y_2 is a prefix of an element of I.

By definition of $|p|$, py_1 cannot be a proper suffix of an element of $C \setminus I$. On the other hand, py_1 has no suffixes in $C \setminus I$ (otherwise, $py_1 = p_1v$ with $v \in C \setminus I$, $p_1 \in A^*$. Since py_1 is a prefix of pt'_n, there exists a $p_2 \in A^*$ such that

$$c't_n = pt'_n = py_1p_2 = p_1vp_2$$

and vp_2 is a suffix of $c't_n$ having $v \in C \setminus I$ as a prefix, contradicting (ii).) Then, by (b), either py_1 is a suffix of an element of I or py_1 has a suffix in I implying

(b') either y_1 is suffix of an element of I or y_1 has a suffix in I.

By (a') and (b'), there is an I-interpretation passing through the point (y_1, y_2) of t'_n. Since I is biprefix and by the arbitrariness of (y_1, y_2), $(t'_n)_{n\in\mathbb{N}}$ is an infinite set of full paths with respect to I. By Lemma 8.3, this is absurd since I is insufficient. □

Lemma 8.8. *D is a complete prefix code.*

Proof. Let us prove that D is a prefix code. Let us suppose the contrary. Then there are $u \in D$, $v \in A^+$ such that $uv \in D$. Then we have two cases:

(1) $u \in I$;

(2) $u = u_1u_2$ with $u_1 \in A^+$, $u_2 \in C\setminus I$.

In the first case, $uv \notin I$ (since $I \subseteq C$ and C is a prefix code) and this contradicts the definition of D since uv is a path constructed by the algorithm (of Lemma 8.4) and it has a prefix in I.

In the second case, $uv = u_1u_2v \in D$ with u_1, $v \in A^+$, $u_2 \in C\setminus I$ and this contradicts Lemma 8.5. Then D is prefix. To prove that D is a complete prefix code, let c be a nonempty path. We must prove that

(*) either c has a prefix in D or c is a prefix of an element of D.

If c has a prefix in I or c is a prefix of an element of I, (*) holds. Otherwise, by Lemma 8.7, there are $t \in A^*$ such that $ct \in A^*$ and a proper suffix z of ct satisfying

(**) either z has a prefix in $C\setminus I$ or z is a prefix of an element of $C\setminus I$.

On the other hand, since c has no prefix in I and c is not a prefix of an element of I, this is true also for ct, i.e., condition (i) of Lemma 8.6 holds for ct.

By (**) and Lemma 8.6 applied to ct, either ct has a prefix in D or ct is a prefix of an element of D. Then (*) holds for c. □

In the following, for any set $X \subseteq A^*$, S_X is the set of the proper suffixes of the elements of X.

Lemma 8.9. *If C has degree d, then D has degree $d+1$.*

Proof. First we note that

$$S_C \subseteq S_D \tag{18}$$

(since $I \subseteq D$ and any element of $C\setminus I$ is a proper suffix of an element of D). Let $c \in C\setminus I$. Then $c \notin H(C)$ (otherwise, $c \in \mathring{C} \subseteq I$) and $c \notin H(D)$ (by Lemma 8.5). By Proposition 5.4(ii) we must prove that c has $d+1$ D-interpretations. By that proposition, c has d C-interpretations and, by Proposition 5.3, there exist $s_1, \ldots, s_d \in S_C$ such that each of them is a prefix of c.

By (18), $s_1, \ldots, s_d \in S_D$. Moreover, $c \in S_D$ (since any element of $C\setminus I$ is a proper suffix of an element of D) and $c \notin S_C$ (since $c \in C$ and C is suffix). If we prove that any $s_{d+1} \in S_D\setminus\{s_1, \ldots, s_d\}$ cannot be a proper prefix of c, then $s_1, \ldots, s_d$, c are the only prefixes of c in S_D. Thus, by Proposition 5.3, c has $d+1$ D-interpretations.

By contradiction, suppose that $s_{d+1} \in S_D\setminus\{s_1, \ldots, s_d\}$ is a proper prefix of c. Since C has degree d, $s_{d+1} \notin S_C$. Thus, s_{d+1} is not a proper suffix of any element of I. By

definition of D, there are u, $t \in A^+$ and $y \in C \setminus I$ such that $us_{d+1} = ty \in D$. Since $s_{d+1} \notin S_C$, we have $|y| \leqslant |s_{d+1}|$.

On the other hand, s_{d+1} is a proper prefix of c; then y is a proper prefix or a proper factor of c. Since $c \in C \setminus I$, c is a proper suffix of an element of D. Then $y \in H(D)$ and this is a contradiction by Lemma 8.5. □

Lemma 8.10. *I is the kernel of D.*

Proof. (1) Let us prove that $\mathring{C} \subseteq \mathring{D} \subseteq I \subseteq C$. Let $c \in \mathring{C}$. Then $c \in D$ since $\mathring{C} \subseteq I \subseteq D$. On the other hand, there are u, $v \in A^+$ such that $ucv \in C$. By definition of D, either ucv is in D or ucv is a proper suffix of an element of D. Thus, $c \in \mathring{D}$. Let $c \in \mathring{D}$. Then $c \in D$ and, by Lemma 8.5, $c \in I$. Thus (1) holds. We must prove that $I = \mathring{D}$. By contradiction, suppose that $I \setminus \mathring{D} \neq \emptyset$.

(2) $\mathring{D}$ is insufficient since $\mathring{D} \subseteq I$. Let D_1 be the finite complete biprefix code obtained by Lemma 8.4 if we take $\mathring{D}$ instead of I. By Lemma 8.9, D_1 has the same degree as D.

(3) Let $v \in I \setminus \mathring{D}$. We have $v \notin H(D)$ (otherwise, $v \in I \subseteq D$ implies $v \in \mathring{D}$, contradicting the hypothesis that $v \in I \setminus \mathring{D}$). Moreover, $v \notin H(D_1)$ (otherwise, $v \in I \setminus \mathring{D} \subseteq C \setminus \mathring{D}$ and $v \in H(D_1)$ contradicts Lemma 8.5 applied to D_1 and $\mathring{D}$ instead of D and I). We show that

(a) $v \in S_{D_1} \setminus S_D$;

(b) $v_1 \in S_D$, v_1 proper prefix of $v \Rightarrow v_1 \in S_{D_1}$;

i.e., a contradiction since, by (2) and Proposition 5.3, the number $d + 1 \geqslant 2$ of prefixes of v in S_D must be equal to the number of prefixes of v in S_{D_1}.

(a): We have $v \in S_{D_1}$ (since $v \in I \setminus \mathring{D} \subseteq C \setminus \mathring{D}$) and $v \notin S_D$ (since $v \in I \subseteq D$ and D is suffix).

(b): Let v_1 be a proper prefix of v such that $v_1 \in S_D$. Then there is a $z \in A^+$ such that either $zv_1 \in I$ or $zv_1 = wy$ with $y \in C \setminus I$ and $w \in A^+$. In the first case, since $\mathring{D} \subseteq D_1$ and $I \setminus \mathring{D} \subseteq C \setminus \mathring{D}$, $v_1 \in S_{D_1}$. In the second case, v_1 is a suffix of y (otherwise, y is a proper factor of $v \in I \setminus \mathring{D} \subseteq D$ with $y \in C \setminus I$, contradicting Lemma 8.5) and we have $v_1 \in S_{D_1}$ since $y \in S_{D_1}$ (by $C \setminus I \subseteq C \setminus \mathring{D}$ and definition of D_1). □

Lemma 8.11. *C is the derivative of D.*

Proof. By definition, the derivative D' of D includes the kernel of D, i.e., by Lemma 8.10, $I \subseteq D'$. The statement can be proven by showing that

$$C \setminus I \subseteq D' \tag{19}$$

(in fact, in this case, we have $C \subseteq D'$; since C, D' are two finite complete biprefix codes, $C \subseteq D'$ implies $C = D'$). Let us prove (19). Let $c \in C \setminus I$. We must prove that $c \in P_D \cap S_D \setminus H(D)$. By Lemma 8.5, $c \notin H(D)$ and, by definition of D, we have $c \in S_D$. Since D is a complete prefix code, there are $u \in D$, $t \in A^+$ such that either

(1) $c = ut$, or

(2) $ct = u$.

In the second case, $c \in P_D$. Moreover, the first case cannot happen. In fact, in case (1), $u \notin I$ ($c \in C \setminus I$ and C is prefix). By definition of D, there exist $u_1 \in A^+$, $u_2 \in C \setminus I$ such that $c = u_1 u_2 t$. Since $c \in S_D$, there exists $u_3 \in A^+$ such that $u_3 c = u_3 u_1 u_2 t \in D$ with $u_3 u_1 \in A^+$, $t \in A^+$, $u_2 \in C \setminus I$, contradicting Lemma 8.5. □

Finally, by Lemmas 8.8, 8.9, 8.10 and 8.11, we have the following main theorem.

Theorem 8.12. *Let C be a finite complete biprefix code of degree d, with kernel $\mathring{C}$, and I an insufficient set such that $\mathring{C} \subseteq I \subseteq C$. Then there exists a finite complete biprefix code D of degree $d+1$ such that $D' = C$ and $\mathring{D} = I$.*

For example, if we take the code

$$C = \{aa, abc, b, abd, dc, ca, cbc, dd, cbd\}$$

on the graph G from Fig. 1 and we take $I = \mathring{C} = \{b\}$, then I is an insufficient set and we have the finite complete biprefix code of degree 3:

$$D = \{b, aaa, caa, aabc, cabc, aabd, cabd, abdc, cbdc, dca, abca, ddc, cbca, dcbc, abcbc, cbcbc, ddd, abdd, cbdd, dcbd, abcbd, cbcbd\}.$$

Acknowledgment

I wish to thank Christophe Reutenauer, for many discussions and suggestions given during this work.

I am also indebted to Dominique Perrin who has read this paper.

This research was partially supported by C.N.R. (Consiglio Nazionale delle Ricerche, Roma, Italy).

References

[1] J. Berstel and D. Perrin, *Theory of Codes* (Academic Press, New York, 1985).
[2] J. Berstel and C. Reutenauer, *Rational Series and Their Languages* (Springer, Berlin, to appear).
[3] Y. Césari, Sur un algorithme donnant les codes bipréfixes finis, *Math. Systems Theory* **6**(3) (1972) 221-225.
[4] Y. Césari, Propriétés combinatoires des codes bipréfixes complets finis, in: D. Perrin, ed., *Actes de la 7ème Ecole de Printemps d'Informatique Théorique* (Jougne, 1979) 29-46.
[5] G. Lallement, *Semigroups and Combinatorial Applications* (Wiley and Sons, New York, 1979).
[6] D. Perrin, Completing biprefix codes, *Theoret. Comput. Sci.* **28** (1984) 329-336.
[7] C. Reutenauer, Intégration des codes bipréfixes (d'après Césari), in: *Séminaire d'Informatique Théorique LITP* (1981-1982) 67-81.
[8] C. Reutenauer, Ensembles libres de chemins dans un graphe, *Bull. Soc. Math. France* **114** (1986) 135-152.
[9] M.P. Schützenberger, On a special class of recurrent events, *Ann. Math. Statist.* **32** (1961) 1201-1213.
[10] B. Tilson, Semigroupoids, to appear.